Renewable and Sustainable Energy Reviews 24 (2013) 103–110

Contents lists available at SciVerse ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

A review of electric load classification in smart grid environment

Kai-le Zhou [a,b,1], Shan-lin Yang [a,b,*], Chao Shen [a,b]

[a] School of Management, Hefei University of Technology, Hefei 230009, China
[b] Key Laboratory of Process Optimization and Intelligent Decision-Making of Ministry of Education, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Article history:
Received 8 January 2013
Received in revised form
25 February 2013
Accepted 4 March 2013
Available online 10 April 2013

Keywords:
Load classification
Smart grid
Process model
Clustering methods and result evaluation methods
Load classification applications

ABSTRACT

The load data in smart grid contains a lot of valuable knowledge, which is useful for both electricity producers and consumers. Load classification is an important issue in load data mining. A five-stage process model of load classification is constructed based on the summary and analysis of studies about load classification in smart grid environment. Then, the commonly used clustering methods for load classification are summarized and briefly reviewed, and the well-known evaluation methods for load classification are also introduced. Besides, the applications of load classification, including bad data identification and correction, load forecasting and tariff setting, are discussed. Finally, an example of load classification based on Fuzzy c-means (FCM) is presented.

Contents

1. Introduction

There is a great difference in the electricity consumption patterns of different types of users, such as domestic, commercial, industrial, agricultural, etc. Even for the same type of users, their patterns of electricity consumption may be different. Mining the electricity consumption patterns of different electricity users based on load data classification can not only support the production

* Corresponding author at: School of Management, Hefei University of Technology, Hefei 230009, Anhui Province, China. Tel.: +86 551 62901008.
[1] Tel.: +86 15955112340.
E-mail addresses: kailezhou@gmail.com, zhoukaile@mail.hfut.edu.cn. (K.-l. Zhou). hgdysl@gmail.com.(S.-l. Yang).

1364-0321/$ - see front matter
http://dx.doi.org/10.1016/j.rser.2013.03.023

Nomenclature

U	membership degree matrix
V	cluster center matrix
n	number of data object
c	number of clusters
μ_{ij}	membership degree of x_j to v_i
E	sum of squared errors of all sample objects
m	fuzziness parameter in FCM
C_i	ith cluster
x_j	jth data object
v_i	cluster center of C_i
$d(x_j,v_i)$	Euclidean distance of x_j to v_i, $d(x_j,v_i)=\|x_j-v_i\|^2$
$J_m(U,V)$	objective function of FCM

Subscripts

i	ith cluster
j	jth data object
k	kth cluster

planning, the making of competitive market policies and the provision of more personalized electric power services for electricity producers [1,2], but also help different electricity users to enhance the understanding of their electricity consumption patterns. Moreover, users can adjust their electricity consumption strategies more economically and optimally based on the knowledge discovered from load classification. Hence, the electricity consumption costs will be reduced and the energy use efficiency will be improved more significantly [3].

Load classification is to partition various load patterns into groups so that load patterns in the same group are more similar to each other than to load patterns in other groups based on various clustering algorithms [4,5]. The characteristic load pattern is used to represent and describe the load patterns in the same group. Load classification is an important part of load modeling, therefore, the accuracy of load classification can directly affect the reasonableness and effectiveness of load modeling [6].

Load classification is a process which consists of many steps. Such as load classification preparation, load classification implementation using clustering method, as well as the understanding and applications of load classification. The process model and specific steps of load classification are presented in Section 2.

While in smart grid environment [7], a large amount of load data will be measured and collected by advanced load measuring equipment. The scale of the load data collected will be larger, and the structure will be more complex. Moreover, the form of load data in smart grid environment will be more flexible. Therefore, mining and extracting valuable knowledge from the massive electric power load data in smart grid environment is an important research direction.

Based on the summation and analysis of existing research about electric power load classification, the five-stage process model of load classification in smart grid environment is established in Section 2. The commonly used clustering methods and result evaluation methods of load classification are reviewed and summarized in Section 3. Section 4 presents the applications of load classification, including bad data identification and correction, load forecasting, and tariff setting, etc. Section 5 gives an example of load classification based on Fuzzy c-means (FCM) algorithm [8]. Finally, conclusions are made in Section 6.

2. Process model of load classification

The electric power load data in smart grid is big. Specifically, its scale is large, its structure is complex and heterogeneous, its dimension is high, and its form is real-time and dynamic. These characteristics make the load classification in smart grid even more difficult. Hence, a definite model of load classification is necessary.

As it is shown in Fig. 1, there are five stages in the process of load classification, namely, load data preparation, load data clustering preparation, load data clustering implementation, understanding and evaluation of load classification results, and applications of load classification results.

The preparation of input data for load classification is the first step. According to the dimensions of time, regions and the types of substations, the power load conditions are determined first. Then selecting sample data from massive load data using sampling methods. Afterwards, the input sample load data selected are normalized, and the outlier and noise data should also be identified and corrected.

Load data clustering preparation stage includes determining the classification characteristics, choosing appropriate clustering algorithm and determining its corresponding parameters. Three kinds of load characteristic indices, descriptive, comparative and curved, are summarized in [9]. The well-known clustering algorithms used for load classification are K-means, FCM, hierarchical clustering method, etc. The parameters in clustering algorithm are, for example, the initial cluster centers, number of clusters, and fuzziness parameter in FCM, etc.

The third step is to implement clustering algorithm based on the pre-processed load data, selected classification characteristics and clustering algorithm and its corresponding parameters.

After the load data clustering, we need to understand and evaluate the classification results. The classification results are generally presented as a certain number of groups of load patterns

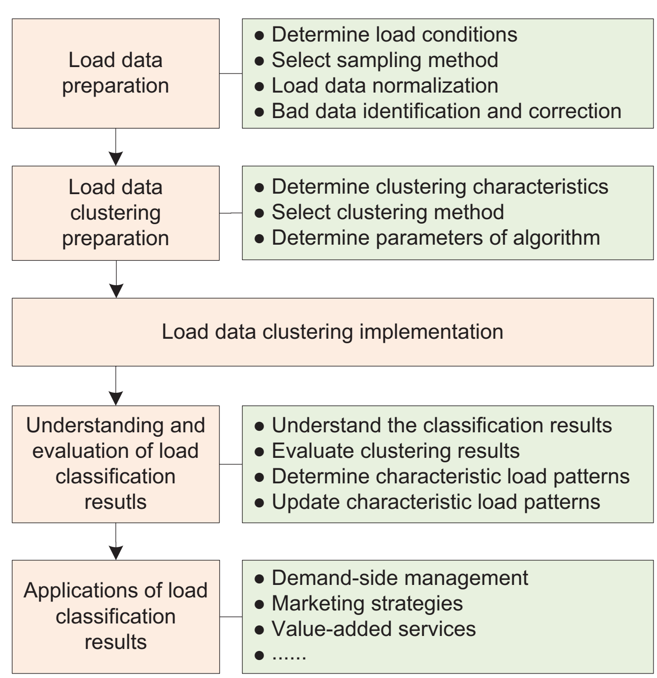


Fig. 1. Process model of load classification.

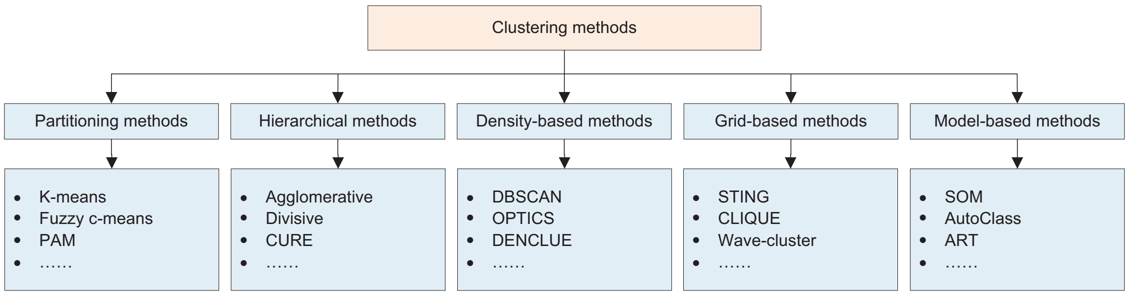


Fig. 2. Clustering methods can be used for load classification.

Table 1
Studies about four commonly used clustering methods for load classification.

Clustering algorithm	Short description	Pros and cons	References
K-means	A classical partitioning crisp clustering method	Simple, efficient and scalable; difficult to determine initial cluster centers and cluster numbers, sensitive to noise and outliers, can only be used to spherical data, etc.	[15–18]
FCM	A well-known local search fuzzy clustering method, also a partitioning method	Membership degree of fuzzy partitions is introduced; difficult to determine initial cluster centers and cluster numbers, easy to fall into local optimum, etc.	[19–24]
Hierarchical clustering method	The bottom-up aggregation or top-down spit of groups	Easy to implement; difficult to select the agglomerate and split points, etc.	[25–28]
SOM	A kind of unsupervised neural networks method	Can identify the most significant characteristics with self-stability, has a strong ability of anti-noise; The learning efficiency depends on the input order of sample objects when the number of objects is small. Be affected by factors such as the weights of network connection, the adjustment of learning efficiency, the selection of neighborhood function, etc.	[29–32]

and their corresponding representative load patterns. The information and characteristics of each group of load patterns need to be described and understood. In addition, cluster validity indices are generally used to validate the quality of clustering results.

The ultimate goal of load classification process is to support the decision-making of power systems participants. Based on the knowledge and information discovered from load classification, the demand-side management can be implemented. Also, it can improve the practicality of bad data identification and correction, the accuracy of load forecasting, and the appropriateness of tariff setting.

3. Clustering methods and result evaluation methods for load classification

3.1. Clustering methods for load classification

Clustering methods [10,11] can be grouped into five categories based on the clustering criterion, and each category contains many specific clustering methods. Such as partitioning methods include K-means, FCM, PAM, etc. All of the clustering methods can be used for load classification, which are summarized in Fig. 2.

We should note that no one clustering method is always superior to the others when they are used for load classification, as they are used for other applications. Some methods are more commonly used for load classification than the others since they are easier to operate or better results can be obtained by them.

We will give a brief introduction to the four commonly used clustering methods for load classification, K-means [12], FCM [8], hierarchical clustering method [13] and self-organization mapping (SOM) [14], from Sections 3.1.1–3.1.4. The four methods and corresponding references are summarized in Table 1.

In addition to the above four commonly used methods, some new methods, such as Support Vector Clustering [33], FaiNet [34], honey bee mating optimization [35], ant colony optimization algorithm [36], fellow the leader [37], iterative refinement clustering [38], and ISODATA [39], etc. have also been studied and used for load classification.

Although there are some differences in the configuration of platforms, software, and hardware when different clustering methods are used for load classification. The key requirements, such as load data measuring and collection platform AMR (Automated Meter Reading), the computing software MATLAB, SPSS or R, and the high-performance computers, are all needed for the implementation of load classification.

3.1.1. K-means algorithm

K-means algorithm [12] is a kind of classical crisp clustering method used for load classification. The basic idea of K-means is that selecting c initial cluster centers randomly once the number of clusters c is determined, then allocate other objects to their nearest cluster according the distance between the object and the cluster centers. Performing iterative operations until the criterion function shown in Eq. (1) converges to a certain range.

$$E = \sum_{i=1}^{c} \sum_{x_j \in C_i} d(x_j, v_i) \tag{1}$$

The operation of K-means is simple, efficient and scalable. Hence, it is the commonly used crisp clustering methods for load classification. However, the deficiencies of K-means include: (1) the selection of initial cluster centers can significantly affect the

Table 2
Well-known CVIs for the evaluation of load classification results.

Type	CVI	References
Crisp clustering	$\text{Dunn} = \min_{i=1,\cdots,c}\left(\min_{\substack{j=1,\cdots,c\\ j\neq i}}\left(\frac{D(C_i,C_j)}{\max_{l=1,\cdots,c}\delta(C_l)}\right)\right)$, where $D(C_i,C_j)=\min_{x\in C_i,y\in C_j} d(x,y), \delta(C_i)=\max_{x,y\in C_i} d(x,y)$	[43]
	6*3 variants of *Dunn* based on different $D(C_i,C_j)$ and $\delta(C_i)$	[44]
	$C=\frac{S(C)-S_{\min}(C)}{S_{\max}(C)-S_{\min}(C)}$, where $S(C)=\sum_{C_k\in C}\sum_{x_i,x_j\in C_k} d(x_i,x_j)$, $S_{\min}(C)=\sum\min(n_w)_{x_i,x_j\in X} d(x_i,x_j)$, $S_{\max}(C)=\sum\max(n_w)_{x_i,x_j\in X} d(x_i,x_j)$	[45]
	$\text{Gamma}=\frac{(S+)-(S-)}{(S+)+(S-)}$	[46]
	$\text{CS}=\frac{\sum_{i=1}^{c}\left(\frac{1}{n_i}\sum_{x_j\in C_i}\max_{x_l\in C_i} d(x_j,x_l)\right)}{\sum_{i=1}^{c}\left(\max_{k\neq i} d(v_i,v_k)\right)}$	[47]
	$\text{I}=\left(\frac{1}{c}\times\frac{E_1}{E_c}\times\max_{i,j=1,\cdots,c} d(v_i,v_j)\right)^p$, where $E_c=\sum_{i=1}^{c}\sum_{x_j\in C_i} d(x_j,v_i)$	[48]
	$\text{SF}=1-\frac{1}{e^{e^{\text{bcd}(C)+\text{wcd}(C)}}}$, where $\text{bcd}(C)=\frac{\sum_{i=1}^{c} n_i\times d(v_i,\overline{X})}{n\times c}$, $\text{wcd}(C)=\sum_{i=1}^{c} 1/n_i\sum_{x_j\in C_i} d(x_j,v_i)$	[49]
	$\text{COP}=\frac{1}{n}\sum_{i=1}^{c} n_i\frac{1/n_i\sum_{x_j\in C_i} d(x_j,v_i)}{\min_{x_i\notin C_i}\max_{x_j\in C_i} d(x_i,x_j)}$	[50]
Fuzzy clustering	$\text{PC}=\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2$, $\text{PE}=-\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}\log_2\mu_{ij}$	[51]
	$\text{FS}=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m(d(x_j,v_i)-d(v_i,\overline{X}))$	[52]
	$\text{XB}=\frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d(x_j,v_i)}{n\times\min_{i\neq j} d(v_i,v_j)}$	[53]
	$\text{NFI}=\frac{c\left(\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2\right)-n}{n(c-1)}$	[54]
	$\text{SC}=\sum_{i=1}^{c}\frac{\sum_{j=1}^{n}\mu_{ij}^m d(x_j,v_i)}{\sum_{j=1}^{n}\mu_{ij}\sum_{k=1}^{c} d(v_i,v_k)}$	[55]
	$\text{PBMF}=\left(\frac{1}{c}\times\frac{E_1}{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d(x_j,v_i)}\times D_c\right)$, where $E_i=\sum_{j=1}^{n}\mu_{ij} d(x_j,v_i)$, $D_c=\max_{i,j=1}^{c} d(v_i,v_j)$	[56]
	$\text{PCAES}=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2/\mu_M-\sum_{i=1}^{c}\exp(-\min_{k\neq i}(d(v_i,v_k)/\beta_T))$, where $\mu_M=\min_{1\le i\le c}\sum_{j=1}^{n}\mu_{ij}^2$, $\beta_T=\frac{1}{c}\sum_{i=1}^{c} d(v_i,\overline{X})$	[57]
	$\text{CO}=C(c,U)-O(c,U)$, $CO_r=\frac{1}{n}\frac{C(c,U)}{O(c,U)}$	[58]

algorithm, (2) determining the appropriate number of clusters is difficult, (3) it is sensitive to noise and outliers data, (4) it can only be used to find groups in spherical data set, etc. Therefore, the K-means algorithm used in load classification is usually modified or optimized [15–18].

3.1.2. FCM

FCM [8] is a well-known local search fuzzy clustering method. A data object in a data set belongs and only belongs to one group in crisp clustering. While in the fuzzy clustering, each data object cannot be strictly clustered into a certain group, but into more than one groups with a certain degree of membership to each group.

FCM algorithm starts with determining the number of clusters followed by guessing the initial cluster centers. Then every data object is assigned a membership degree to each cluster. Each cluster center point and corresponding membership degree are updated iteratively by minimizing the objective functions until the positions of the cluster centers does not change or the difference of objective function values between two iterations ranges in a permitted extent.

The objective function of FCM algorithm is defined as

$$J_m(U,V)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d(x_j,v_i) \tag{2}$$

The iterative procedure updates membership μ_{ij} and the cluster centers v_i by

$$\mu_{ij}=\frac{1}{\sum_{k=1}^{c}(d(x_j,v_i)/d(x_j,v_k))^{1/m-1}} \tag{3}$$

$$v_i=\frac{\sum_{j=1}^{n}\mu_{ij}^m x_j}{\sum_{j=1}^{n}\mu_{ij}^m} \tag{4}$$

where μ_{ij} satisfies

$$\mu_{ij}\in[0,1] \tag{5}$$

$$\sum_{i=1}^{c}\mu_{ij}=1,\forall j=1,\ldots,n \tag{6}$$

$$0<\sum_{j=1}^{n}\mu_{ij}<n,\forall i=1,\ldots,c \tag{7}$$

The concept of membership degree of fuzzy partitions is introduced to FCM, and it has become a popular fuzzy clustering method for load classification [19–24]. However, FCM is also sensitive to the initial cluster centers, and the cluster number is also difficult to be determined. Moreover, it is easy to fall into local optimum. All of these factors can affect the accuracy and effectiveness of load classification. The research of FCM optimized by intelligent algorithm is an interesting direction [24].

3.1.3. Hierarchical clustering method

The main idea of hierarchical method [13] is the bottom-up aggregation or top-down spit of groups in a data set until the satisfied classification result is formed. Each object in the data set is regarded as a group, and then form a larger one by merging two groups based on a certain criterion (generally the distances among clusters) until all the objects are in a single cluster or meeting a termination condition in the agglomerate-type hierarchical

clustering. The steps of split-type hierarchical clustering are just opposite to the agglomerate-type.

The steps of hierarchical method are easy to implement, thereby it is widely used for load classification. However, the selection of agglomerate and split points is difficult, since the clustering operation is based on the former steps, and the former steps cannot be changed, the final clustering result is directly affected by the merging or split operation in each step. Hence, the hierarchical clustering method is also improved and modified when used for load classification [25–28].

3.1.4. Self-organization mapping (SOM)

SOM [14] is a kind of unsupervised neural networks method, also known as Kohonen neural network. SOM is composed of input layer and competitive layer. There are N input neurons in the input layer and M output neurons in the output layer. The neurons in input layer and output layer are interconnected, and the winning neuron is the nearest one in output layer to N input neurons.

Evaluation function is not needed in SOM, and it can identify the most significant characteristics with self-stability. SOM also has a strong ability of anti-noise. All of these make SOM being widely used for load classification [29–32]. However, the learning efficiency depends on the input order of sample objects when the number of objects is small. The factors such as the weights of network connection, the adjustment of learning efficiency, the selection of neighborhood function, can significantly affect the performance of SOM, thereby affect the effectiveness of load classification.

3.2. Evaluation methods for load classification results

Since clustering is an unsupervised process, the load data objects in data sets are unlabeled and no structural knowledge about the data set is available. Hence, measuring the quality of clustering results and determining the optimal number of clusters are difficult tasks. The most commonly used approach to determine the optimal number of cluster is to execute the clustering algorithm several times with different number of clusters and then selecting the number of clusters that provides the best result observing a predefined criterion function. The predefined criterion function is called cluster validity index (CVI). When the number of clusters and other parameters of clustering algorithm are fixed, CVI can be used to evaluate and validate the results of load classification. Currently, a large number of CVIs have been proposed and reviewed [40–42]. Previous studies on CVIs have demonstrated that there is no single CVI that can deal with any data sets and always perform better than the others. But they are consistent on the basic principle that a good partition should have a small intra-cluster variance and a large inter-cluster separation at the same time. Here, we review some well-known CVIs which can be used for the evaluation of load classification results and present a brief summary shown in Table 2.

4. Applications of load classification

4.1. Bad data identification and correction based on load classification

The bad data existing in the load data set can affect the correct decision-making of power producers, and even affect the daily running and the safety of power systems [59]. In smart grid environment, power producers and managers must accurately identify and appropriately process the bad load data effectively.

Many studies have focused on bad data identification and correction based on load classification. Zhang et al. [60] presented an intelligent cleaning model for bad data based on load classification using Kohonen neutral network optimized by fuzzy soft clustering. While Wang et al. [61] identified the bad data effectively using K-means clustering algorithm based on cluster validity index, thereby reducing the undetected and false detected bad data. Similarly, the bad data in transmission grid state estimation were detected, identified and corrected by K-means algorithm combining validity index in [62]. Additionally, Jiang et al. [63] identified the bad data according to the good data classification obtain by fuzzy equivalent matrix clustering.

Existing studies have demonstrated that the effectiveness and practicality of bad data identification and correction can be improved by load classification. The results obtained by load classification are the input of bad data identification and correction, and it is an important influencing factor.

4.2. Load forecasting based on load classification

Load forecasting is a hot research direction in demand-side management of power systems, especially in smart grid environment, and various load forecasting methods have been proposed [64].

Load classification can also be used for load forecasting, and the accuracy of load forecasting can be improved supported by load classification. Misiti et al. [65] grouped the global electric power information based on clustering methods, and then obtained the overall forecasting results by combining the decomposed forecasting information. While Li and Han [66] presented a load forecasting method based on ant colony clustering, which can improve the accuracy of load forecasting. Also, Jota et al. [28] gave a load forecasting method of daily load curves and the peak load based on the typical daily curve and the corresponding dynamic load model obtained by hierarchical clustering.

Load classification can be used to support the accurate forecasting of load in many ways. Such as forecasting the total load based on the load of different types of users obtained from load classification. In addition, each type of consumers' typical load profile or characteristic load can be used as the input data of load forecasting.

4.3. Tariff setting based on load classification

Many countries are taking deregulation and open marketing polices of electricity market in smart grid environment. These polices can promote the competitions in electricity market, improve the efficiency of investments and power systems operation, and reduce costs. In China, the tariff reform has become the

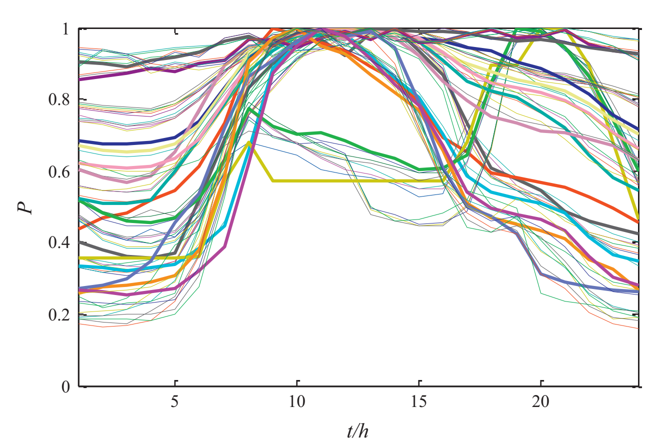


Fig. 3. 72 load profiles.

core of the power systems reform. Developing differentiated and personalized tariff according to the load constitution of distribution network and the consumption patterns of different users is an important and significant research area [67].

Mahmoudi et al. [68] pointed out the knowledge of how and when consumers use electricity is essential to the retailer in competitive environment, and proposed an annual framework for optimal price offering by a retailer based on the clustering and classification of load profiles of consumers. While Chicco et al. [69] analyzed the tariff setting and costs of power distribution companies based on the classification of consumers' load profiles. Huang et al. [70] presented a tariff decision-making model by considering load classification and electricity using characteristics from load rate, the power supply voltage level, the load shape and the reliability requirements, etc. Also, Ozveren et al. [71] proposed a method for the automatic classification of large-sets of electrical demand profiles using fuzzy relation, and the classification results can be used by Supply Business for tariff development and end user costing.

The tariff setting is a complex process and various techniques, such as optimization, decision-making, and economics, are required.

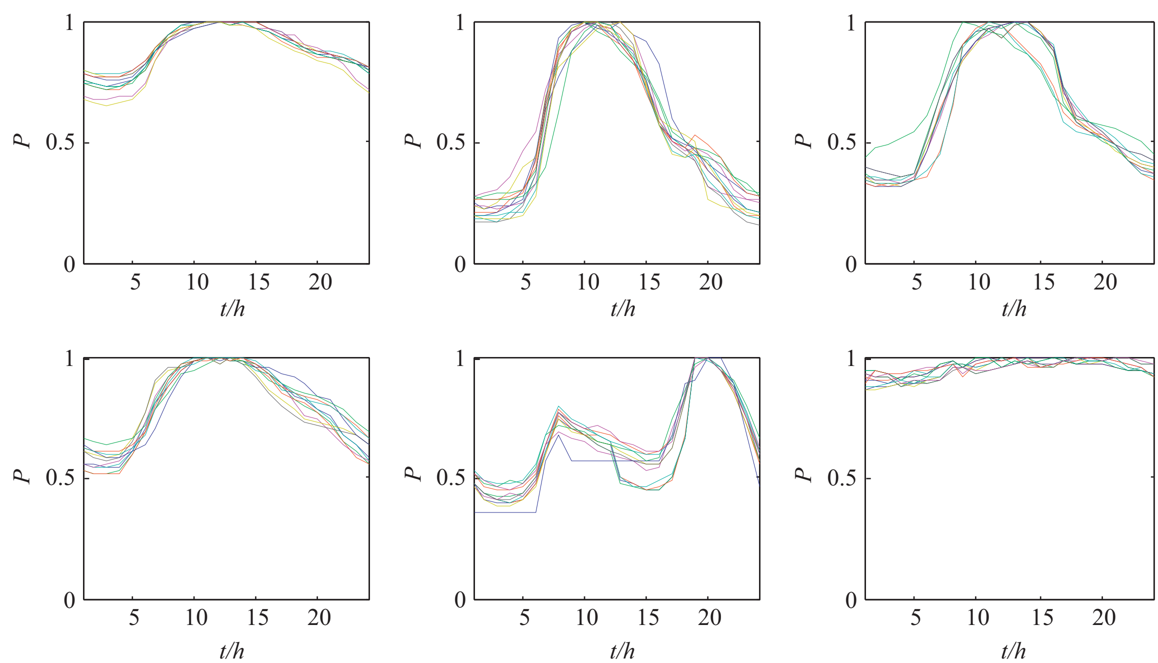


Fig. 4. Load profiles in each group.

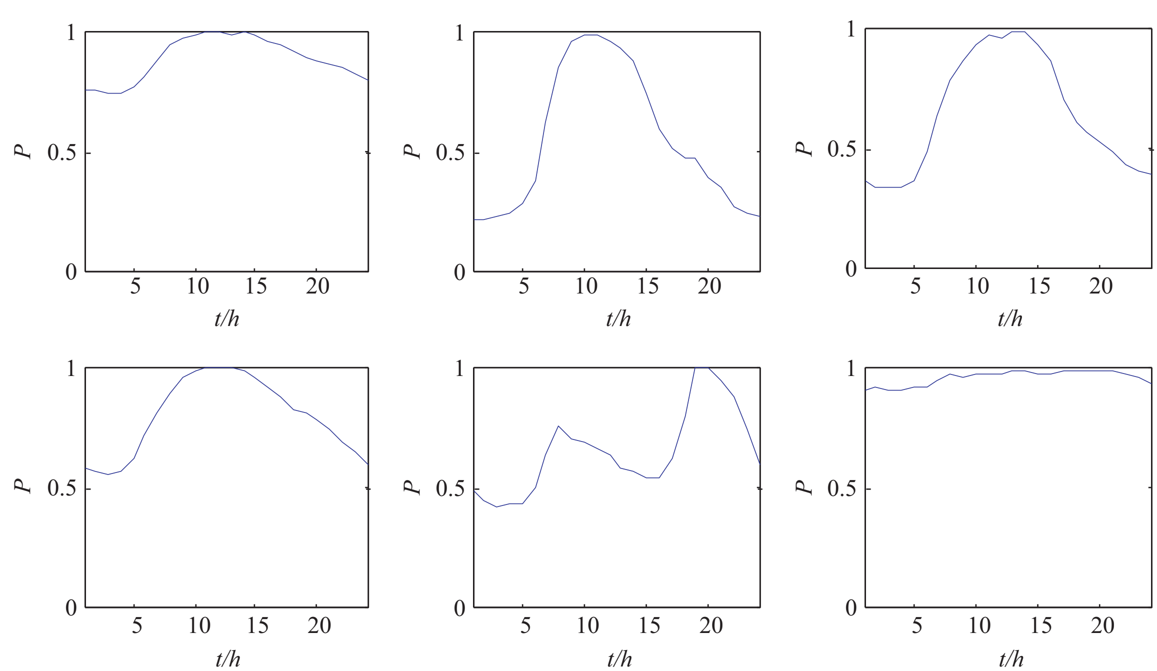


Fig. 5. Characteristic load profile of each group.

Load classification is an important decision support tool for tariff setting.

5. An example of load classification based on FCM

In order to illustrate the process of load classification, we present an example based on FCM algorithm. The data used are 72 load profiles of 6 types of different electricity consumers in a city of China [72], each load profile is a daily load profile measured every one hour. The load profiles are shown in Fig. 3.

The parameters in FCM are set as follows. The number of clusters $c=6$, the fuzziness parameter $m=2.5$, the initial cluster centers are selected randomly, and the algorithm is implement 50 times and the average values are selected as the result. According to the process model described in Section 2, the load classification results are shown in Figs. 4 and 5.

As Figs. 4 and 5 show, the shape of load profiles, which indicate the pattern of electricity consumption of different users, are different. For example, the range of load profiles in the second group is about 0.2–1.0, which is a larger range. While the entire load values in the sixth group are high, with smaller range. Also, different from the other five groups of load profiles, there are two peaks in the fifth group, one of which appears in the night. Additionally, more information and knowledge can be discovered from the results of load classification. Based on these, the decision-making and policies development can be more effective and efficient for both electricity producers and consumers.

6. Conclusion

With the in-depth theoretical study and widespread application of smart grid, load classification will play an increasingly important role in decision-making of power systems and service provision of electricity market. Load classification methods are the premise and basis of load classification, and the analysis and applications are the ultimate goal of load classification. The difficulties and research directions of load classification in smart grid are as follows.

(1) The influence of the complex smart grid environment to load classification. The load data in smart grid environment are massive, dynamic, high-dimensional, and heterogeneous. All these characteristics increase the difficulty of load classification in each process. Such as the efficient update of characteristic load patterns with the adding and deleting of consumers.
(2) The study of efficient and effective load classification methods. Traditional clustering algorithms, such as K-means, FCM and hierarchical method, are widely used, but the deficiencies of these methods have been demonstrated, which can significantly affect the effectiveness of load classification. Moreover, most traditional clustering methods are inefficient in dealing with the big load data in smart grid. Hence, more efficient new methods and the optimized traditional methods should be developed. While evaluating and validating the load classification results, we should not only consider the values of CVIs, but also the characteristics of load data and the purpose of load classification.
(3) The study of before-classification preparation and after-classification analysis. In addition to the bad data identification and processing, data normalization and the selection of load classification characteristics, data sampling and reduction methods are also important research contents in the preparation process of load classification. While in the after-classification process, more effective and efficient methods about the evaluation, understanding and analysis of classification results need to be studied.
(4) The expansion study of load classification in demand-side management [73]. Besides bad data identification and correction, load forecasting and tariff setting based on load classification, there are more applications based on the results of load classification, which are interesting research directions.

Acknowledgements

This work is supported by the National High Technology Research and Development Program (863 Program) of China (No. 2011AA05A116) and the National Natural Science Foundation of China (No. 71131002 and No. 71071045).

References

[1] Sathiracheewin S, Surapatana V. Daily typical load clustering of residential consumers. In: Eighth international conference on electrical engineering/electronics, computer, telecommunications and international technology (ECTI-CON); 2011. p. 797–800.
[2] Valero S, Ortiz M, Senabre C, et al. Methods for customer and demand response policies selection in new electricity markets. IET Generation, Transmission & Distribution 2007;1(1):104–10.
[3] Grandjean A, Adnot J, Binet G. A review and an analysis of the residential electric load curve models. Renewable and Sustainable Energy Reviews 2012;16(9):6539–65.
[4] Bidoki SM, Mahmoudi-Kohan N, Gerami S. Comparison of several clustering methods in the case of electrical load curves classification. In: 16th conference on electrical power distribution networks (EPDC); 2011. p. 1–7.
[5] Chicco G, Napoli R, Piglione F. Application of clustering algorithms and self organising maps to classify electricity consumers. In: IEEE Bologna power tech conference proceedings. IEEE; 2003. p. 23–26.
[6] Lin SJ, Li XR, Liu YH, et al. A classification method for aggregated load dynamic characteristics. Automation of Electric Power Systems 2005;29(22):33–8.
[7] Wu YN, Chen J, Liu LR. Construction of China's smart grid information system analysis. Renewable and Sustainable Energy Reviews 2011;15(9):4236–41.
[8] Bezdek JC, Ehrlish R, Full W. FCM: the fuzzy c-means clustering algorithm. Computers and Geosciences 1984;10(2–3):191–203.
[9] Zhao XX, Zhou XQ, Jiang SJ. Analysis and forecast of China's electricity load characteristics. China Electric Power Press 2011.
[10] Jain A, Dubes R. Algorithms for clustering data. Prentice Hall. 1988.
[11] Xu R, Wunsch II D. Survey of clustering algorithms. IEEE Transactions on Neural Networks 2005;16(3):645–78.
[12] Macqueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability; 1967. p. 281–97.
[13] Johnson JC. Hierarchical clustering schemes. Psychometrika 1967;32(3):241–54.
[14] Kohonen T. The self-organization maps. Proceedings of the IEEE 1990;78(9):1464–80.
[15] Kim YI, Shin JH, Song JJ, et al. Customer clustering and TDLP (typical daily load profile) generation using the clustering algorithm. In: Asia and Pacific transmission & distribution conference & exposition 2009; 10:1–4.
[16] Liu L, Wang G, Zhai DH. Application of k-means clustering algorithm in load curve classification. Power System Protection and Control 2011;23(65–68):7339 2011;23(65–68):73.
[17] Räsänen T, Voukantsis D, Niska H, et al. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Applied Energy 2010;87(11):3538–45.
[18] López JJ, Aguado JA, Martín F, et al. Hopfield–K-means clustering algorithm: a proposal for the segmentation of electricity consumers. Electric Power System Research 2011;81(2):716–24.
[19] Ju P, Jin Y, Wu F, et al. Studies on classification and synthesis of composite dynamic loads. Automation of Electric Power Systems 2004;28(1):64–8.
[20] Li XR, Jiang XJ, Qian J, et al. A classifying and synthesizing method of power consumer industry based on the daily load profile. Automation of Electric Power Systems 2010;34(10):56–61.
[21] Zokaria Z, Lo KL. Two-stage fuzzy clustering approach for load profiling. In: Proceedings of the 44th international universities power engineering conference (UPEC) 2009; 9: p. 1–5.
[22] Li PQ, Li XR, Chen HH, et al. The characteristics classification and synthesis of power load based on fuzzy clustering. In: Proceedings of the Csee; 2005. 25(24): p. 73–78.
[23] Yang H, Zhang L, He Q, et al. Study of power load classification based on adaptive fuzzy C means. Power System Protection and Control 2010;16(111–115):2238 2010;16(111–115):22.

[24] Zhou KL, Yang SL. An improved fuzzy c-means algorithm for power load characteristics classification. Power System Protection and Control 2012;40 (22):58–63.
[25] Gerbec D, Gasperic S, Gubina F. Determination and allocation of typical load profiles to the eligible consumers. In: IEEE Bologna power tech conference proceedings. IEEE; 2003. p. 23–26.
[26] Kim YI, Ko JM, Choi SH. Methods for generating TLPs (typical load profiles) for smart grid-based energy programs. In: IEEE symposium on computational intelligence applications in smart grid(CIASG), IEEE; 2011. p. 1–6.
[27] Li XR, Lin SJ, Liu YH, et al. A new classification method for aggregate load dynamic characteristics based on field measured response. In: Proceedings of the Csee; 2006. 26(8): p. 39–44.
[28] Jota PRS, Silva VRB, Jota FG. Building load management using cluster and statistical analyses. International Journal of Electrical Power & Energy Systems 2011;33(8):1498–505.
[29] Verdu SU, Garcia MO, Franco FJG, et al. Characterization and identification of electrical consumers through the use of self-organizing maps and daily load parameters. IEEE power systems conference and exposition. IEEE; 2004. p. 899–906.
[30] Gavrilas M, Ivanov O, Gavrilas G. Load profiling with fuzzy self-organizing algorithms. In: Ninth symposium on neural network application in electrical engineering (NEUREL); 2008. p. 25–27.
[31] Li ZY, Wu JY, Wu WL, et al. Power consumers load profile clustering using the SOM neural network. Automation of Electric Power Systems 2008;15(66–70):832 2008;15(66–70):8.
[32] Zhang HB, He RM, Liu YM. The characteristics clustering and synthesis of electric dynamic loads based on Kohonen neural network. In: Proceedings of the Csee; 2003. 23(5): p. 1–5,43.
[33] Chicco G, Llie IS. Support vector clustering of electrical load pattern data. IEEE Transactions on Power Systems 2009;24(3):1619–28.
[34] Gu DZ, Ai Q, Chen C. The application of artificial immune network in load classification. In: Third international conference on electric unility deregulation and restructuring and power technologies; 2008. p. 1394–1398.
[35] Gavrilas M, Gavrilas G, Sfintes CV. Application of honey bee mating optimization algorithm to load profile clustering. In: IEEE international conference on computational intelligence for measurement systems and applications (CIMSA). IEEE; 2010. p. 113–118.
[36] Sun YM, Wang CL, Zhang ZS, et al. Clustering analysis of power system load series based on ant colony optimization algorithm. In: Proceedings of the Csee; 2005. 25(18): p. 40–45.
[37] Chicco G, Napoli R, Piglione F, et al. Load pattern-based classification of electricity consumers. IEEE Transactions on Power Systems 2004;19(2):1232–9.
[38] Batrinu F, Chicco G, Napoli R, et al. Efficient iterative refinement clustering for electricity customer classification. IEEE Russia Power Technology IEEE 2005:1–7.
[39] Mutanen A, Ruska M, Repo S, et al. Customer classification and load profiling method for distribution systems. IEEE Transactions on Power Delivery 2011;26 (3):1755–63.
[40] Milligan G, Cooper M. An examination of procedures for determining the number of clusters in a data set. Psychometrika 1985;50(2):159–79.
[41] Kim M, Ramakrishna RS. New indices for cluster validity assessment. Pattern Recognition Letters 2005;26(15):2353–63.
[42] Wang W, Zhang Y. On fuzzy cluster validity indices. Fuzzy Sets and Systems 2007;158(19):2095–117.
[43] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics 1973;3:32–57.
[44] Bezdek JC, Pal NR. Some new indexes of cluster validity. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 1998;28:301–15.
[45] Hubert L, Schultz J. Quadratic assignment as a general data-analysis strategy. British Journal of Mathematical and Statistical Psychology 1976;29:190–241.
[46] Baker FB, Hubert LJ. Measuring the power of hierarchical cluster analysis. Journal of the American Statistical Association 1975;70:31–8.
[47] Chou HC, Su MC, Lai E. A new cluster validity measure and its application to image compression. Pattern Analysis and Applications 2004;7:205–20.
[48] Bandyopadhyay S, Maulik U. Nonparametric genetic clustering: comparison of validity indices. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 2001;31(1):120–5.
[49] Saitta S, Raphael B, Smith I. A bounded index for cluster validity. Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science 2007;4571:174–87.
[50] Gurrutxaga I, Albisua I, Arbelaitz O, et al. SEP/COP: an efficient method to find the best partition in hierarchical clustering based on a new cluster validity index. Pattern Recognition 2010;43:3364–73.
[51] Bezdek JC. Numerical taxonomy with fuzzy sets. Journal of Mathematical Biology 1974;7(1):57–71.
[52] Fukuyama Y, Sugeno M. A new method of choosing the number of cluster for the fuzzy c-means method. In: Proceedings of the fifth fuzzy systems symposium; 1989. p. 247–250.
[53] Xie XL, Beni G. A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 1991;13(8):841–7.
[54] Pal NR, Bezdek JC. On cluster validity for fuzzy c-means model. IEEE Transactions on Fuzzy Systems 1995;3(3):370–9.
[55] Bensaid AM, Hall LO, Bezdek JC, et al. Validity-guided (re)clustering with applications to image segmentation. IEEE Transactions on Fuzzy Systems 1996;4(2):112–23.
[56] Pakhira MK, Bandyopadhyay S, Maulik U. A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification. Fuzzy Sets and Systems 2005;155:191–214.
[57] Wu KL, Yang MS. A cluster validity index for fuzzy clustering. Pattern Recognition Letters 2005;26(9):1275–91.
[58] Žalik KR. Cluster validity index for estimation of fuzzy clusters of different sizes and densities. Pattern Recognition 2010;43(10):3374–90.
[59] Liu L, Zhai DH, Jiang XL. Current situation and development of the methods on bad-data detection and identification of power system. Power System Protection and Control 2010;38(5):143–52.
[60] Zhang XX, Cheng QY, Zhou Q, et al. Dynamic intelligent cleaning for dirty electric load data based on data mining. Automation of Electric Power Systems 2010;38(5):143–52.
[61] Wang BS, Duan ZQ, Zhai DH. Application of k-means algorithm based on effective index to identifying bad data of the power system. Northeast Electric Power Technology 2010;3(16–18):31.
[62] Liu L, Wang G, Zhai DH. Application of k-means clustering algorithm in load curve classification. Power System Protection and Control 2011;23(65–68):7339 2011;23(65–68):73.
[63] Jiang DL, Wang KW, Wang XD. Clustering method of fuzzy equivalence matrix to bad-data detection and identification. Power System Protection and Control 2011;21(1–6):1139 2011;21(1–6):11.
[64] Beccali M, Cellura M, Lo Brano V, et al. Short-term prediction of household electricity consumption: assessing weather sensitivity in a Mediterranean area. Renewable and Sustainable Energy Reviews 2008;12(8):2040–65.
[65] Misiti M, Misiti Y, Oppenheim G, et al. Optimized clusters for disaggregated electricity load forecasting. Statistical Journal 2010;8(2):105–24.
[66] Li W, Han ZH. Short-term power load forecasting using improved ant colony clustering. In: First international workshop on knowledge discovery and data mining; 2008. p. 221–224.
[67] Chua SC, Oh TH, Goh WW. Feed-in tariff outlook in Malaysia. Renewable and Sustainable Energy Reviews 2011;15(1):705–12.
[68] Mahmoudi-Kohan N, Moghaddam MP, Sheikh-El-Eslami MK. An annual framework for clustering-based pricing for an electricity retailer. Electric Power Systems Research 2010;80(9):1042–8.
[69] Chicco G, Napoli R, Postolache P, et al. Customer characterization options for improving the tariff offer. IEEE Transactions on Power Systems 2003;18 (1):381–7.
[70] Huang YH, Kang CQ, Xia Q, et al. Study on consumer classification differentiated pricing of electricity. Electric Power 2004;37(1):24–8.
[71] Ozveren CS, Vechakanjana C, Birch AP. Fuzzy classification of electrical load demand profiles-a case study. In: fifth international conference on power system management and control; 2002. p. 353–358.
[72] Zhang ZH. Study on load classification in power system. Thesis. Tianjin University; 2007.
[73] Kale RV, Pohekar SD. Electricity demand supply analysis: current status and future prospects for Maharashtra, India. Renewable and Sustainable Energy Reviews 2012;16(6):3960–6.